IJPHCS
Open Access: e-Journal

International Journal of Public Health and Clinical Sciences
e-ISSN : 2289-7577. Vol. 5:No. 1
January/February 2018

THE FACTORS THAT CONTRIBUTE TO DIABETES MELLITUS IN MALAYSIA: ALTERNATIVE LINEAR REGRESSION MODEL APPROACH IN THE HEALTH FIELD INVOLVING DIABETES MELLITUS DATA

Awang Nawi M. A.[1*], Ahmad W. M. A. W.[2], Mamat M.[1]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA).
[2]School of Dental Science, Universiti Sains Malaysia (USM)

**Corresponding author: Mohamad Arif Awang Nawi, Faculty Informatics and Computing, Universiti Sultan Zainal Abidin, UniSZA 21300 Kuala Terengganu, Terengganu, Malaysia, Email: anz_wsh@yahoo.com*

ABSTRACT

Background: Diabetes mellitus (or diabetes) is a common disease that can cause of morbidity and mortality. Besides that, it is a serious deadly disease that making someone very weak and infirm. In Malaysia, the World Health Organization (WHO) has estimated the number of 0.94 millions of diabetics in 2000 will increase by 164% in the next 30 years which means the total number of diabetics is 2.48 millions in 2030. This study aimed to obtain significant factors associated with diabetes mellitus among patients in Malaysia.

Materials and Methods: The study also focused on efficiency model between multiple linear regression and alternative linear regression based on R-Square, adj R-Square, significant risk factors (p-value) and average width of the interval coefficients for each independent variable. Multiple linear regression and alternative linear regression analysis were used to identify risk factors contribute to diabetes mellitus among patients in Malaysia. The accepted level of significance was set below 0.05 ($p<0.05$) and all these methods are improved the programming language by using SAS 9.3 software.

Result: From the linear regression model, there is only one variable that contributes to diabetes mellitus among patients that is high factor ($\beta = 12.82526$, $p < 0.0351$). Whereas, by using alternative linear regression analysis, all independent variables such as mass index ($\beta = -4.44754$, $p < 0.0001$), total cholesterol ($\beta = 0.06689$, $p < 0.0001$), height ($\beta = -1.98315$, $p < 0.0001$), systolic blood pressure ($\beta = 0.06941$, $p < 0.0001$) and the weight (lbs) ($\beta = 0.79864$, $p < 0.0001$) are significant to diabetes mellitus. Average width of former multiple regression was found to be 61188.298 while using alternative linear regression model, the average width are 118.019.

Discussion and Conclusion: From this analysis, the most efficient method of obtained relationship between response and explanatory variable is alternative linear regression method compared to linear regression method.

Keywords: Diabetes Mellitus, Alternative Linear Regression Model and Multiple Linear Regressions

International Journal of Public Health and Clinical Sciences
e-ISSN : 2289-7577. Vol. 5:No. 1
January/February 2018

1.0 Introduction

In Malaysia, there are many diseases that are caused by Diabetes mellitus (DM). For example, premature and preventable mortality, heart disease, end stage renal failure, blindness, stroke, and amputation. The burden of diabetes continues to increase and being a major public health concern in Malaysia when The National Health and Morbidity Survey (NHMS) 2011 has shown that diabetes in Malaysia has increased from 11.6% in 2006 to 15.2 % in 2011. This means that within five years, the percentage of diabetes has increased by 3.2 %. This study aimed to obtain significant factors associated with diabetes mellitus among patients in Malaysia. The study also focused on efficiency model between multiple linear regressions and alternative linear regression based on R-Square, adjusted R-Square, significant risk factors (p-value) and the average width of the interval coefficients for each independent variable (Institute for Public Health (IPH), 2011).

In the whole world, there are 135 millions of people were estimated had diabetes in 1995, while 154 million of diabetics in the year of 2000 and it is expected to top 300 million in the year of 2025 with the main increase being in the developing countries. In the developed countries, the estimated increased are 42% meanwhile in the developing countries is 170% (King et al., 1998).

Diabetes mellitus (or diabetes) is a common disease that can cause morbidity and mortality. Besides that, it is a serious deadly disease that making someone very weak and infirm. This disease has now reached to epidemic proportions and the prevalence rates are expected to go even higher in the foreseeable future. In Malaysia, the Fist National Health and Morbidity Survey (Institute for Public Health (IPH), 2006) was conducted in 1986 and Second National Health and Morbidity Survey (King et al., 1998) were conducted in 1996. In NHMS 1 reported that the prevalence of DM is 6.3%. Meanwhile, the reading show increasing to 8.2% in NHMS II (Public Health Institute Ministry of Health Malaysia, 1997). In Malaysia, the World Health Organization (WHO) has estimated the number of 0.94 millions of diabetics in 2000 will increase by 164% in the next 30 years which means the total number of diabetics is 2.48 million in 2030.

Obesity has become a global problem, especially in developing countries and in developed countries. One of the risks faced by obese people is diabetes because is strongly related to obesity and prevalence increase of diabetes is because of a rise in the prevalence of obesity (Astrup and Finer, 2000). In a study in Kelantan, 38.4% of diabetics were either obese or overweight compared to 24.1% in those with normal glucose tolerance (Mafauzy et al., 1999). In 1997, a diabetes management conducts a study in major government hospitals in Malaysia. The study showed that many of the patients had not been given enough intensive care. Only 30% had urine albumin checked, 22% had blood lipids measured and 10% had Alemeasured (Mustaffa et al., 1998).

The general practitioners in Peninsular Malaysia conduct a study in 2001 that showed the majority of patients were not well controlled with a high prevalence of complications. Only 20% had Alcof < 7%, 12.3% had total cholesterol of < 4.8 moll and 44.1% had systolic blood pressure of < 140 mmHg. Neuropathy was the most common complication (30.1%), followed by background retinopathy (23.5%), albuminuria (22.9%) and microalbuminuria (20.4%) (Mafauzy, 2005).

International Journal of Public Health and Clinical Sciences
e-ISSN : 2289-7577. Vol. 5:No. 1
January/February 2018

2.0 Materials and Method

Multiple linear regression and alternative linear regression analysis were used to identify risk factors contribute to diabetes mellitus among patients in Malaysia. Table 1 shows the description of dependent and independent variables. The accepted level of significance was set below 0.05 ($p<0.05$). Multicollinearity was checked between independent variables. Alternative linear regression method is a combination of four main methods which are linear regression, robust regression, bootstrap and fuzzy linear regression. And, all these methods are improved the programming language by using SAS 9.3 software (Ahmad et al., 2016).

Table 1: Description of the Dependent and Independent Variable for Diabetes Study

Variable in Diabetes Data		
Diabetes	Y	Diabetes mellitus among patients in Malaysia
BMI	X_1	Body mass index diabetes patients
Cholesterol	X_2	The total of cholesterol in the patient's body
Height (cm)	X_3	High patient
Systolic	X_4	Systolic blood pressure patients
Weight (lbs)	X_5	weight patients

The multiple linear regression modelsis expressedas $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \ell$ where β's are crisp parameters and X_n are vector of crisp numbers. Usually, ℓ are assumed to be independent random variables with mean of 0 and variance σ^2. The parameters are usually estimated using method of least squares. The explanation of various aspects of multiple linear regression methodology is given in Hoffmann Draper and Smith (Draper and Smith, 1998).

Fuzzy numbers which can be presented by interval, for example, A_i can be express as fuzzy set given by $A_i =< a_{1c}, a_{1w} >$ where a_{ic} is centre and a_{iw} is radius or vagueness associated. Fuzzy setabove reflects theconfidencein theregression coefficients around a_{ic} in terms of symmetric triangular memberships function. Application of this method should be given more attention when the underlying phenomenon is fuzzy which means that the response variable is fuzzy. So, the relationship is also considered to be fuzzy. This $A_i =< a_{1c}, a_{1w} >$ can be written as $A_1 = [a_{1L}, a_{1R}]$with $a_{1L} = a_{1c} - a_{1w}$ and $a_{1R} = a_{1c} - a_{1w}$[12]. In fuzzy linear regression methodology, parameters are estimated by minimizing total vagueness in the model.

$$y_j = A_0 + A_1 x_{1j} + A_2 x_{2j} + \ldots + A_k x_{kj}$$

Using $A_i =< a_{1c}, a_{1w} >$ we can write

$$y_j =< a_{0c}, a_{0w} > + < a_{1c}, a_{1w} > x_{1j} + \ldots + < a_{nc}, a_{nw} > x_{nj}$$
$$=< a_{jc}, a_{jw} > .$$

Thus $y_{jc} = a_{0c} + a_{1c}x_{1j} + \cdots + a_{nc}x_{nj}$

$$y_{jw} = a_{0w} + a_{1w}\left|x_{1j}\right| + \cdots + a_{nw}\left|x_{nj}\right|$$

As y_{jw} represent radius and so cannot be negative, therefore on the right-hand side of equation $y_{jw} = a_{0w} + a_{1w}\left|x_{1j}\right| + \cdots + a_{nw}\left|x_{nj}\right|$, absolute values of x_{ij} are taken. Suppose there m data point, each comprising $a(n+1) - row$ vector. Then parameters A_i are estimated by minimizing the quantity, which is total vagueness of the model-data set combination, subject to the constraint that each data point must fall within estimated value of response variable. Figure 1 shows the steps of forming an alternative linear regression model to obtain significant factors associated with diabetes mellitus among patients in Malaysia.

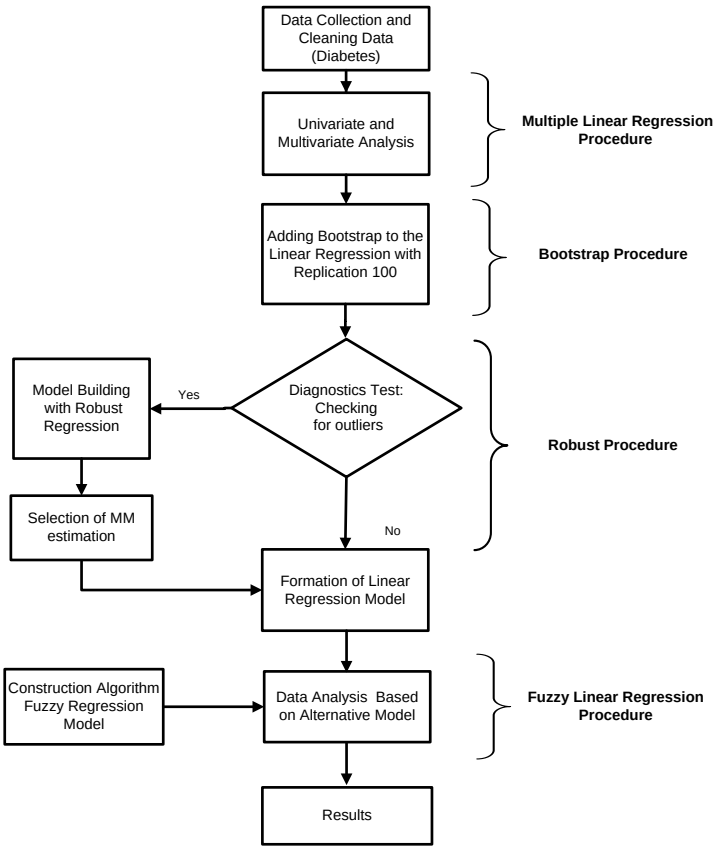

Figure 1: Flow Chart of Alternative Linear Regression Model

3.0 Results

Table 2 shows the factors associated with diabetes mellitus among patients in Malaysia by using multiple linear regressions. In the model below there is only one variable that contributes to diabetes mellitus among patients that is a high factor (β = 12.82526, p<0.0351).

Whereas, four of variables such as body mass index, total cholesterol, systolic blood pressure and the weight (lbs) is not the significant factor in the development of diabetes among patients.

Table 2: Factors Associated with Diabetes Mellitus among Patients in Malaysia by Using Multiple Linear Regressions

Independent variable	DF	β	SE	p-Value
BMI	1	34.51741	17.26388	0.0570
Cholesterol	1	0.14858	0.19701	0.4581
Height (cm)	1	12.82526	5.74296	0.0351*
Systolic	1	0.65496	0.43126	0.1419
Weight (lbs)	1	-5.37402	2.63026	0.0522
R-Square	0.2770			
Adj R-Sq	0.1264			
Average Width	61188.298			

DF denotes the degree of freedom.
B denotes the standardized variable estimate.
SE denotes the standard error of the variable estimate.
*$p< 0.05$

The result from Table 3 is alternative linear regression model which combined with robust regression by using MM estimator, bootstrap technique, and fuzzy linear regression. Based on the significant level (less than < 0.05) and can be concluded that the regression model is acceptable because there are at least one significant independent variables (F (5, 2594) = 942.23, p<.0001). Table 3 showed that all independent variables such as mass index (β = -4.44754, p< 0.0001), total cholesterol (β = 0.06689, p<0.0001), height (β = -1.98315, p< 0.0001), systolic blood pressure (β = 0.06941, p< 0.0001) and the weight (lbs) (β = 0.79864, p< 0.0001) are significant to diabetes mellitus. Based on all the above factors influence and contribute to diabetes disease among patients in Malaysia. R square (R^2) is proportion varies from 0 and 1. As can be seen from Table 3, the value of R^2 is 0.6449, which means that 64.49 percent of the total variance in diabetes mellitus among patients in Malaysia has been 'explained'.

Table 3: Factors Associated with Diabetes Mellitus among Patients in Malaysia by Using Alternative Linear Regression Model Approach (n = 100)

Independent variable	DF	β	SE	p-Value
BMI	1	-4.44754	0.32287	.0001***
Cholesterol	1	0.06689	0.00336	.0001***
Height (cm)	1	-1.98315	0.11068	.0001***
Systolic	1	0.06941	0.00705	.0001***
Weight (lbs)	1	0.79864	0.04946	.0001***
R-Square	0.6449			
Adj R-Sq	0.6442			
Average Width	118.019			

DF denotes the degree of freedom.
B denotes the standardized variable estimate.
SE denotes the standard error of the variable estimate.
***$p< 0.0001$

4.0 Discussion and Conclusion

In this study, we evaluated the independent effect of diabetes mellitus among patients in Malaysia by using multiple linear regression and alternative linear regression model. From the Table 4 below show that comparison between these two methods. The main findings of our study were that all risk factors of diabetes mellitus including body mass index (BMI) (β = -4.69287, p< 0.0001), cholesterol (β = 0.06803, p<0.0001), height (β = -2.13611, p< 0.0001), systolic blood pressure (β = 0.07009, p< 0.0001) and weighty (β = 0.83355, p< 0.0001) of diabetes mellitus were independently associated among patients in Malaysia. Our findings are consistent with previous studies showing that body mass index (BMI), cholesterol, height, systolic blood pressure and weight are associated with diabetes mellitus among patients (Luk et al., 2014).

Otherwise, multiple linear regression models showed that only one risk factor that contributes diabetes mellitus among patients in Malaysia is height (β = 12.825, p< 0.0351). From Table 4, the average width of former multiple regression was found to be 61188.298 while using alternative linear regression model, the average width is 118.019. From this analysis, the most efficient method to an obtained relationship between response and explanatory variable is alternative linear regression method compared to linear regression method. Methodology in this study was supported by the Ahmad et al., 2016 and they stated that an alternative method for multiple linear model regression modeling, a technical combining with robust, bootstrap and fuzzy approach the most efficient method compared to linear regression method (Ahmad et al., 2016).

Table 4: Comparison between Multiple Linear Regression Model and Alternative Linear Regression Model (n = 100)

Model	Multiple Linear Regression Model	Alternative Linear Regression Model (n = 100)
F-Test	(F (5, 24) = 1.84, p> 0.1431)	(F (5, 2594) = 942.23, p< 0.0001)
Goodness of Fit	R^2 = 0.2770 Adj R-Sq = 0.1264	R^2 = 0.6449 Adj R-Sq = 0.6442
Independent Variables	BMI (x_1) (β = 34.517, p> 0.0570) Cholesterol (x_2) (β = 0.149, p>0.4581) Height (x_3) (β = 12.825, p< 0.0351) Systolic (x_4) (β = 0.655, p> 0.1419) Weight (x_5) (β = -5.374, p> 0.0522)	BMI (x_1) (β = -4.44754, p< 0.0001) Cholesterol (x_2) (β = 0.06689, p<0.0001) Height (x_3) (β = -1.98315, p< 0.0001) Systolic (x_4) (β = 0.06941, p< 0.0001) Weight (x_5) (β = 0.79864, p< 0.0001)
Average Width	61188.298	118.019

Limitations of linear regression model often to get the attention of researchers from various fields of study, particularly in the health sciences. Therefore, an alternative method that has been developed to improve multiple regression models known as alternative linear regression

models to approach. The flexibility and accuracy models can be considered as the outliers. Data outliers can spoil and mislead the training process resulting in a longer training time, less accurate models and ultimately unsatisfactory results. This is because the linear regression model cannot detect that there are outliers in the data studied (Mokhtar, 1994). Alternative methods can be applied to data constraints faced by researchers. Besides that, this method can be applied when there is a problem of accuracy in the estimation of statistics, especially in determining the confidence interval. This technique can be used as a measure of uncertainty and bias, especially on the parameters estimation of independent variables and scattered similar. The alternative methods also produce the average width smaller than traditional methods. The requirement of this alternative model has been the impetus for researchers especially in the medical field that involves a small data set (Efron and Tibshyrani, 1993).

Acknowledgments

This research was supported by Ministry of Higher Education in Malaysia for their financial contribution in respect for this study. The author would like to thank our university from Universiti Sultan Zainal Abidin (UniSZA) who provided insight and expertise that greatly assisted the research. The author also is thankful to Mustafa Mamat and Wan Muhamad Amir W Ahmad for the suggestions throughout the research for assistance with the particular technique, methodology and for comments that greatly improved the manuscript. The author would like grateful to Mustafa Mamat, Wan Muhamad Amir W Ahmad for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

References

Ahmad. W.M.A.W., Nawi. M.A.A., Aleng. N.A. & Mohamad. S. 2016. An Alternative MethodforMultiple Linear Model Regression Modeling, a Technical Combining of Robust,Bootstrap andFuzzy Approach. *Journal of Modern Applied Statistical Methods.* Vol. 15:Iss. 2, Article 44.

Ahmad, W.M.A.W., Nawi, M.A.A. & Mamat, M. 2016. A new strategy of handling generalinsurancemodeling using applied linear method. Malaysian *Journal of Applied Sciences*,1(1): 45-54.

Astrup. A. & Finer. N. 2000. Redefining type 2 diabetes: "diabesity" or "obesity dependent diabetesmellitus"? *Obesity Reviews*. 1(2); 57-59.

Draper, N., & Smith, H. (1998). *Applied regression analysis (3rd ed.)*. New York: Wiley.

Efron B. & Tibshyrani R.J. 1993. *"An introduction to the bootstrap",* ed. Chapman and Hall, New-York.

Institute for Public Health (IPH) 2006. *The Third National Health and Morbidity Survey 2006 (NHMSIII)*.Diabetes.

Institute for Public Health (IPH) 2011. *National Health and Morbidity Survey 2011 (NHMS2011).* Vol.II: Non- Communicable Diseases.

Kacprzyk, J., & Fedrizzi, M. 1992. *Fuzzy regression analysis.* Warsaw: Omnitech Press.

King. H, Aubelt. R.E. & Herman. W.H. 1998. Global Burden of Diabetes and Projections, 1995- 2025.*Diabetes Care*. 21: 1414-31.

Luk AOY, Zhang Y, Ko GTC, Brown N &Ozaki R. 2014. Health-Related Quality of Life inChinesePatients with Type 2 Diabetes: *An Analysis of the Joint Asia Diabetes Evaluation (JADE)Program. J Diabetes Metab* 5: 333.

Mafauzy. M., Mokhtar. N., Wan Mohamad. W.B. & Musalmah. M. 1999. Diabetes mellitus and associated cardiovascular risk factors in North East Malaysia. *Asia Pacific Journal of Public Health*. ll (l): 16-19.

Mafauzy. M. 2005. Diabetes control and complications in private primaly healthcare inMalaysia.*Medical Journal of Malaysia*. 60 (2): 212-17. 398.

Mokhtar Abdullah. 1994. *Regression Analysis*. Dewan Bahasa dan Pustaka. Kuala Lumpur.

Mustaffa. B.E., Wan Mohamad. W.B., Chan. S.P., Rokiah. P., Mafauzy. M. & Kumari. S. 1998. Thecurrent status of diabetes management in Malaysia. *The Journal of the ASEANFederation ofEndocrine Societies*. 16(2): 1-13.

Public Health Institute Ministry of Health Malaysia. 1997. *Report of the Second National Health andMorbidity Survey Conference 20-22 November 1997*, Kuala Lumpur.